

Instrumen Penilaian Kemampuan Berpikir Kritis Peserta Didik SMA Negeri Mata Pelajaran Biologi Kelas X

Instrument for Assessing the Critical Thinking Ability of High School Students on Biology Subject X Grade

TAUFIQ SATRIA MUKTI ^{1*}, EDI ISTIYONO²

¹² Progam Pascasarjana, Universitas Negeri Yogyakarta, Jalan Colombo, Karang Malang, Yogyakarta, 55281, Indonesia
*taufiq.satria2016@student.uny.ac.id

Manuscript received: Revision accepted:

ABSTRACT

Research development is done by quantitative approach that aims to: (1) Generate tools for the assessment of critical thinking ability; (2) Describe the quality of instruments developed to measure critical thinking skills. The study population is the XIP MIPA class of fourteen SMAN in Kendal District. The sample of the research was determined using purposive sampling technique which represented the school with high, medium, and low ranking criteria. Trial conducted on 201 learners. The test instrument was developed with a reasonable multiple choice form. Four categories of scores on each item correspond to the scoring of the PCM (Partial Credit Model) politomus model. Research data were analyzed using the help of EXCEL application, SPSS 16, QUEST and PARSCALE. The results showed that: (1) The critical thinking test instrument fulfilled the validity of the content of the expert judgment; (2) The test instrument with four categories of fit scores on the PCM model; (3) The test instrument has a high reliability value; (4) The test instrument has good difficulty level criteria; (5) Test instruments can be used to estimate critical thinking skills.

Keywords: Critical Thinking, Class X Biology, Politomus Scoring, Partial Credit Model, Multiple Choice Grounded

PENDAHULUAN

Penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar (Mardapi, 2016). Tertulis dalam Peraturan Menteri Pendidikan dan Kebudayaan Tahun 2016 Nomor 23 Tentang Standar Penilaian Pendidikan bahwa penilaian bertujuan untuk memantau perkembangan hasil belajar siswa dan mengevaluasi proses pembelajaran. Urgensi dalam melakukan penilaian menurut (Sudjana, 2013: 3 dan Mardapi, 2012: 14) harus didasarkan pada standar operasi kurikulum yang berlaku. Dengan demikian tujuan instruksional dalam pelaksanaan penilaian berjalan secara sistematis.

Pelaksanaan pendidikan dalam mengimplementasikan penilaian kemampuan berpikir kritis secara umum masih sangat rendah, yaitu sekitar 45% (Lane, 2016). Temuan penelitian (Huber, C.H dan Kuncel, N. R. 2016) menunjukkan bahwa dalam pelaksanaan penilaian kemampuan berpikir kritis dalam pembelajaran membutuhkan waktu lebih lama dari pada penilaian standar yang telah ditetapkan. Temuan lain bahwa penelitian yang dilakukan sudah dilakukan dalam mengestimasi kemampuan berpikir kritis umumnya hanya berfokus pada pencapaian nilai berdasarkan standar kompetensi dan tes yang dilakukan (Anisa, 2017). Selain itu, tidak banyak peneliti yang meneliti secara mendalam butir maupun instrumen yang digunakan dan juga peneliti belum

mengeksplorasi kemampuan keseluruhan siswa berdasarkan respon.

Penerapan penilaian kemampuan berpikir kritis sering dilakukan dengan menggunakan tes esai. Hal ini sesuai dengan konsep berpikir kritis (Ennis, 1986) bahwa dalam pelaksanaan tes perlu ada keterlibatan mental, strategi dan representasi yang digunakan untuk memecahkan masalah, membuat keputusan, dan mempelajari konsep-konsep baru. Namun, sebenarnya (McPeck, J. 1981) menemukan bahwa bentuk esai terdapat kelemahan dan permasalahan yang tidak dapat ditoleransi dalam proses penilaian, yaitu munculnya efek subjektivitas pada pemeriksaan hasil tes.

Kemampuan berpikir kritis jarang diukur dengan menggunakan model tes pilihan ganda. Ini terjadi karena terdapat banyak faktor guessing dalam implementasi (Stephen, 1988) dan membutuhkan keahlian khusus dalam membuat item tes. Namun demikian (Hartini, 2015) dan (Akbar et al, 2017) menemukan bahwa keterampilan berpikir kritis dapat diukur dengan menggunakan tes pilihan ganda, yaitu butir pilihan ganda yang mngedepankan kemampuan HOT (High Order Thinking). Sedangkan (Wilson, 2006) dan (McPeck, J. 1981) menambahkan bahwa instrumen tes yang digunakan dalam mengukur kemampuan berpikir kritis harus memiliki tingkat kesulitan yang tinggi.

Pentingnya mengukur kemampuan berpikir kritis menurut pendapat (Travis, 2015) bahwa berpikir kritis merupakan kemampuan esensial yang dapat digunakan sebagai indikator keberhasilan belajar dalam mencapai

standar kompetensi. Selain itu, tes yang digunakan mengukur kemampuan berpikir kritis dalam pelajaran Biologi juga merupakan bentuk pelatihan dalam menghadapi dan menyelesaikan permasalahan dalam kehidupan nyata (Palm, 2008) dan juga sejalan dengan konsep pembelajaran sains (Towle, 1989: 16-31) yang selalu mengedepankan pemikiran kritis untuk dapat memahami setiap pelajaran yang sangat dekat dengan objek nyata.

Kendala dalam mengestimasi kemampuan berpikir kritis siswa dalam bentuk tes esai dan pilihan ganda membutuhkan solusi lain dalam bentuk tes dengan model baru. MCR (Multiple-Choice With Reason) adalah bentuk tes dengan memilih jawaban dan memilih alasan tertutup. Bentuk tes ini telah terlebih dulu dikembangkan (Istiyono, 2014) yang diasumsikan dapat mengestimasi dan menggambarkan kemampuan berpikir kritis karena melibatkan proses berfikir dan pengetahuan yang ada untuk menyelesaikan masalah yang kompleks. Aspek berpikir kritis yang dikembangkan dalam indikator tes MCR meliputi asumsi, argumen, analisis, evaluasi, dan kesimpulan.

Artikel penelitian yang ditulis ini akan membahas tentang konstruksi tes berpikir kritis berdasarkan data empiris. Analisis data menggunakan kecocokan model berdasarkan IRT (Item Respon Theory) untuk melihat item yang sesuai untuk digunakan berdasarkan kriteria yang ada. Tujuan lain dalam penulisan artikel ini adalah untuk mendeskripsikan kualitas instrumen tes yang meliputi kecocokan model, reliabilitas tes, tingkat kesulitan, dan fungsi informasi instrumen tes.

METODE PENELITIAN

Mengadopsi dari Metode pengembangan tes Mardapi (2016) dalam penelitian pengembangan instrumen tes berpikir kritis ini dilakukan dengan langkah yang meliputi: 1) Menyusun spesifikasi tes. 2) Menulis tes. 3) Meninjau validitas isi. 4) Melakukan uji coba 5) Menganalisis item tes.

Unsur pokok dalam menyusun spesifikasi tes adalah bentuk tes dan panjang tes/lama pelaksanaan tes. Penelitian pengembangan Instrumen tes berpikir kritis yang dilakukan adalah mengadopsi bentuk tes pilihan ganda beralasan yang terlebih dulu dikembangkan Istiyono (2013) dalam mengukur kemampuan HOT pada Pelajaran Fisika. Pada setiap butir soal terdapat empat kategori skor dengan ketentuan : Kategori-1 jika jawaban salah dan alasan salah; Kategori-2 jika jawaban benar dan alasan salah; Kategori-3 jika jawaban salah dan alasan benar; Kategori-4 jika jawaban benar dan alasan benar. Tes berpikir kritis yang dikembangkan sebanyak 45 butir dengan tes A dan B yang masing-masing 25 butir (5 anchor) yang dilaksanakan selama 2 jam pelajaran Biologi sesuai dengan peraturan pemerintah pada jenjang SMA.

Tahap penulisan tes terdiri dari penentuan kompetensi yang akan diujikan sesuai dengan teori berpikir kritis, penentuan materi yang akan di ujikan, penyusunan kisi-kisi, penulisan butir, penyusunan pedoman penskoran,

validitas isi dan perbaikan butir. Kompetensi yang akan diujikan adalah pada mata pelajaran Biologi kelas X MIPA SMA Negeri. Sedangkan teori berpikir kritis yang digunakan meliputi: Aspek 1 (A1): Asumsi, subaspek 1 (SA11): menentukan hipotesis yang relevan, subaspek 2 (SA12): menentukan hasil pertimbangan berdasarkan latar belakang dan fakta; Aspek 2 (A2): Argumentasi, subaspek 1 (SA21): membuat argumen berdasarkan fakta dan pengetahuan, subaspek 2 (SA22): Identifikasi hubungan sebab akibat/alasan; Aspek 3 (A3): Analisis, subaspek 1 (SA31): Analisis latar belakang dan tujuan informasi, subaspek 2 (SA32): mengkaitkan informasi dengan aktivitas manusia; Aspek 4 (A4): Evaluasi, subaspek 1 (SA41): memeriksa kesesuaian masalah dengan solusi, subaspek 2 (SA42): membuat kritik terhadap suatu permasalahan; Aspek 5 (A5): Menyimpulkan, Subaspek 1 (SA51): Menginduksi pemikiran berdasarkan informasi, subaspek 2 (SA52): Mendeduksi pemikiran berdasarkan informasi. Pada tahap penulisan tes juga sangat memerlukan matriks tabel agar instrumen yang dikembangkan sesuai dengan aspek, subaspek dan materi Biologi kelas X SMA semester 1 yang terdiri dari (1) Ruang Lingkup Biologi, 2) Metode Ilmiah, 3) Klasifikasi, 4) Keanekaragaman Hayati, 5) Jamur, 6) Protista, 7) Bakteri, 8) Virus.

Validitas isi dilakukan oleh *expert judgment* yang terdiri dari ahli pengukuran, penilaian dan ahli Pembelajaran Biologi. Pada tahapan ini validitas isi yang dilakukan adalah melihat kesesuaian kompetensi, indikator dan materi Biologi yang telah ditulis menjadi butir soal dalam instrumen tes berpikir kritis. Tujuan dilakukan validitas isi adalah untuk mengetahui kelayakan butir soal yang ditinjau dari segi konsep, konstruksi, bahasa dan keefektifan butir soal untuk mengukur tingkat kemampuan berpikir kritis siswa.

Sebelum dilakukan uji coba instrumen, perlu mengkaji subjek tes yang memenuhi kriteria kemampuan siswa dengan kategori rendah, sedang dan tinggi. Langkah tersebut dilakukan dengan melihat peringkat sekolah berdasarkan Nilai UN tahun 2017 pada mata pelajaran Biologi SMA Negeri. Uji coba dilakukan pada siswa kelas X MIPA SMA Negeri di Kabupaten Kendal yang mewakili dari sekolah dengan peringkat tinggi, sedang, dan rendah yaitu sebanyak 201 siswa. Jumlah tersebut berdasarkan ketemtun analisis berdasarkan IRT (Item Respon Theory) untuk memenuhi analisis 1 PL dengan model PCM.

Pelaksanaan uji coba dilakukan dengan set tes A dan B dengan posisi tempat duduk peserta didik yaitu depan, belakang, kanan, dan kiri mengerjakan soal dengan berselang-seling dengan kode soal A dan B. Hal ini bertujuan untuk mengurangi kecurangan pada saat pelaksanaan uji coba tes. Selain itu, pada pelaksanaan uji coba melibatkan guru mata pelajaran Biologi pada kelas masing-masing untuk menjadi pengawas, dengan tujuan agar pada saat pelaksanaan peserta didik bersungguh-sungguh dalam mengerjakan tes.

Data hasil respon peserta didik pada tahap uji coba yang perlu di kaji adalah berupa kecocokan model (*goodness of fit tes*), tingkat kesukaran, fungsi informasi,

ICC, standar eror pengukuran dan reliabilitas. Pada analisis data diperlukan bantuan program aplikasi Excel, SPSS, Quest, dan Parscale.

HASIL DAN PEMBAHASAN

Instrumen Tes kemampuan berpikir kritis pada mata pelajaran Biologi kelas X SMA Negeri kurikulum 2013 yang dikembangkan terdiri dari 45 butir soal. Instrumen Tes dikembangkan berdasarkan teori Berpikir Kritis yang meliputi aspek dan subaspek serta bersesuaian dengan materi Biologi Kelas X semester 1 kurikulum 2013. Instrumen yang dihasilkan sudah melalui beberapa tahapan sebelum dilakukan uji coba. Beberapa tahapan yang dimaksud adalah penelaah isi, konstruksi, maupun bahasa yang dilakukan oleh ahli pengukuran, penilaian dan praktisi (Guru), sehingga instrumen dinyatakan layak untuk di uji cobakan. Sebaran butir yang dikembangkan adalah pada tabel 1 sebagai berikut:

Tabel 1 : Sebaran Butir Soal Tes Kemampuan Berpikir Kritis

A	SA	Materi Biologi Kelas X Semester 1				
		1	2	3	4	5
A1	SA11		7,8			
	SA12		9,10			
A2	SA21	1,2		13,14		43,44
	SA22					
A3	SA31			15,14		45,22
	SA32	3,4		41,42		25,26
A4	SA41		11,12			
	SA42				15,18	20,21
A5	SA51				16,17	
	SA52				5,6	24,28

A	SA	Materi Biologi Kelas X Semester 1		
		6	7	8
A1	SA11		33,34	
	SA12			
A2	SA21			
	SA22	31,32	35,36	
A3	SA31			
	SA32			37,38
A4	SA41	29,45	39,40	
	SA42			18,19
A5	SA51			
	SA52			

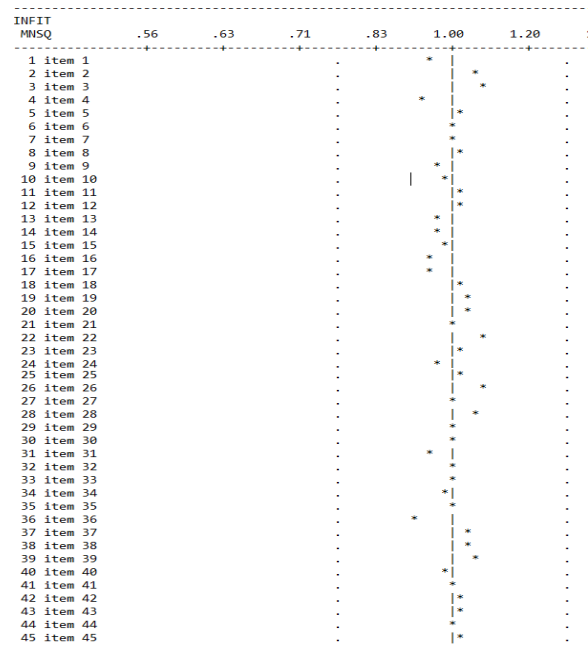
Setiap butir soal terdapat 5 pilihan jawaban dan 5 pilihan alasan. Sehingga penskoran yang dilakukan terhadap butir adalah dengan skor kategori 1-4. Skor 1 jika jawaban salah dan alasan salah; skor 2 jika jawaban benar dan alasan salah; skor 3 jika jawaban salah dan alasan benar; skor 4 jika jawaban dan alasan benar. Berdasarkan pada teori pengukuran skor kategori dapat dilakukan analisis dengan model penskoran politomus (1-4) adalah dengan model PCM (*partial credit model*) 1-PL.

Hasil yang diperoleh berdasarkan penelitian pengembangan instrumen tes berpikir kritis pada tahap uji coba adalah sebagai berikut:

1. Kecocokan Butir Instrumen (*Goodness Of Fit Test*)

Analisis kecocokan model instrumen tes dilakukan dengan program aplikasi QUEST. Beberapa kategori yang menyatakan kecocokan model menurut Adam dan Khoo (1996:30) adalah dengan melihat rata-rata nilai INFIT *Mean Of Square (Mean INFIT MNSQ)* dan nilai simpangan bakunya. Selain itu dapat juga dilakukan dengan melihat nilai rata-rata INFIT *t (Mean INFIT t)* dan simpangan bakunya. Batasan nilai pada keseluruhan butir yang dinyatakan *fit* adalah jika nilai INFIT *Mean Of Square* $\pm 1,00$ dan simpangan bakunya 0,00. Hasil analisis data yang diperoleh menunjukkan bahwa nilai INFIT *Mean Of Square* 1,00 dan nilai simpangan bakunya adalah 0,04, sehingga dapat disimpulkan bahwa keseluruhan butir pada tes *fit* untuk model PCM 1 PL.

Selain kecocokan model instrumen tes, ketentuan lain adalah dengan melihat kecocokan tiap butir soal yang harus dinyatakan *fit* terhadap model. Menurut Adam dan Khoo (1996:30) setiap butir dinyatakan *fit* dan sesuai dengan model jika nilai INFIT MNSQ berada di antara 0,77 sampai dengan 1,30. Berdasarkan diagram pada Gambar 1 menunjukkan bahwa seluruh butir dinyatakan *fit*. Hal tersebut terlihat bahwa seluruh butir berada pada garis batas *goodness of fit* yaitu pada rentang nilai 0,9 sampai dengan 1,00.



Gambar 1: Diagram INFIT MNSQ

2. Hasil Estimasi

Analisis terhadap hasil uji coba instrumen yang dilakukan dengan model penskoran politomus (PCM 1 PL) diperoleh karakteristik instrumen tes berpikir kritis Mata

Pelajaran Biologi Peserta didik SMA Negeri kelas X semester 1 kurikulum 2013 adalah meliputi hasil estimasi pengukuran, reliabilitas, dan tingkat kesulitan pada tabel 2 sebagai berikut :

Tabel 2: Hasil Estimasi Butir

No.	Uraian	Estimasi Butir	Estimasi Testi
1	Mean and SD	$-0,2 \pm 0,68$	$0,11 \pm 0,18$
2	Mean and SD Adjusted	$0,0 \pm 0,59$	$0,0 \pm 0,11$
3	Reliability	0,74	-
4	Mean and SD Infit MNSQ	$1,01 \pm 0,04$	$1,00 \pm 0,22$
5	Mean and SD Outfit MNSQ	$1,00 \pm 0,09$	$1,00 \pm 0,34$
6	Mean and SD Infit t	$0,08 \pm 0,51$	$0,00 \pm 1,14$
7	Mean and SD Outfit t	$0,05 \pm 0,46$	$0,01 \pm 0,99$

Berdasarkan data yang diperoleh pada tabel 1 dapat disimpulkan bahwa seluruh kriteria yang digunakan dalam mengetahui kecocokan model (*goodness of fit tes*) seluruhnya memenuhi syarat dan berada pada kategori yang baik.

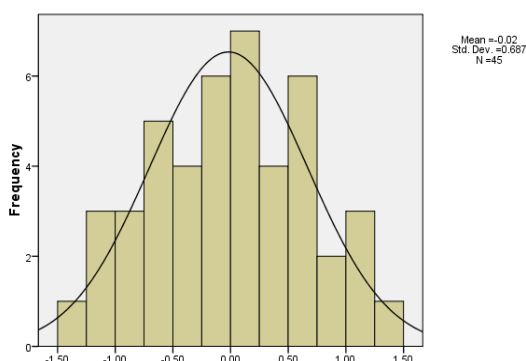
3. Tingkat Kesulitan

Analisis tingkat kesulitan yang dilakukan dengan menggunakan program aplikasi QUEST sesuai dengan model PCM 1-PL diperoleh hasil tingkat kesulitan butir sebagai berikut :

a. Sebaran Tingkat Kesulitan Butir

Hasil analisis terhadap tingkat kesulitan (*difficulty*) seluruh butir berada pada tingkat kesulitan dengan rentang -1,28 sampai 1,39 dengan rata-rata 0,00 dan simpangan baku 0,67. Hasil analisis yang diperoleh ini dapat disimpulkan bahwa seluruh butir berada pada kategori tingkat kesulitan yang baik, karena seluruh butir mempunyai tingkat kesulitan lebih dari -2,00 dan kurang dari 2,00.

Tingkat kesulitan seluruh butir instrumen tes ini dapat divisualisasikan melalui histogram pada Gambar 2. Sebaran tingkat kesulitan 45 butir pada tahap uji coba yang dianalisis dengan menggunakan SPSS mengindikasikan bahwa sebaran tingkat kesulitan mendekati distribusi normal, akan tetapi tidak ada persyaratan bahwa untuk satu tes harus mempunyai tingkat kesulitan dengan sebaran yang normal.



Gambar 2: Sebaran Tingkat Kesulitan Seluruh Item

b. Tingkat Kesulitan Pada Subaspek

Tingkat kesulitan Instrumen tes berpikir kritis yang dikembangkan dengan aspek dan subaspek berpikir kritis berdasarkan hasil analisis uji coba (diperoleh tingkat kesulitan pada tabel 2 sebagai berikut :

Tabel 3: Tingkat Kesulitan Pada Aspek dan Subaspek

A	Rata-rata	SA	Rata-rata
A1	-0.21	SA11	-0.03
		SA12	-0.40
A2	0.17	SA21	-0.04
		SA22	0.37
A3	0.21	SA31	0.34
		SA32	0.09
A4	-0.08	SA41	-0.28
		SA42	0.13
A5	-0.27	SA51	-0.36
		SA52	-0.19

Berdasarkan tabel diatas secara jelas dapat dipahami bahwa rata-rata tingkat kesulitan yang ada adalah dari tingkat kesulitan setiap butir soal pada instrumen tes yang dikembangkan sesuai dengan mariks yang kisi-kisi disusun. Adapun matriks tersebut adalah berupa kesesuaian antara aspek, subaspek dan materi Biologi.

c. Tingkat Kesulitan Tertinggi dan Terendah Butir

Hasil analisis tingkat kesulitan butir tes kemampuan berpikir kritis dapat disajikan butir termudah dan tersulit pada tabel 4 sebagai berikut:

Tabel 4: Butir dengan Tingkat Kesulitan Tertinggi dan Terendah

No	Tertinggi		Terendah	
	Butir	Tk. Kesulitan	Butir	Tk. Kesulitan
1	45	1.16	10	-1.33
2	44	1.1	15	-1.15
3	43	0.96	1	-1.09
4	42	1.06	7	-1.06
5	23	1.34	17	-0.99

Butir dengan tingkat kesulitan tertinggi sebesar 1.34 adalah pada butir 23 sedangkan butir dengan tingkat kesulitan terendah sebesar -1.33 adalah pada butir 10.

d. Presentase Aspek Dan Subaspek Tiap Kategori Skor Butir

Hasil analisis yang diperoleh berdasarkan persentase respon butir dengan kategori skor 1, 2, 3, dan 4 sesuai pada aspek dan subaspek instrumen tes adalah pada tabel 5 sebagai berikut :

Tabel 5: Persentase Respon Jawaban Dengan Kategori Skor

A	SA	Kategori Skor (%)			
		1	2	3	4
A1	SA11	33.60	20.05	6.88	39.45
	SA12	19.25	8.40	16.03	56.18

A2	SA21	38.58	10.84	4.26	46.26
	SA22	49.80	21.58	6.90	21.75
A3	SA31	40.46	13.28	20.60	25.66
	SA32	35.12	13.50	4.70	46.70
A4	SA41	19.28	22.53	2.50	55.73
	SA42	34.82	17.10	10.68	37.42
A5	SA51	26.15	8.98	5.58	59.35
	SA52	37.00	12.20	11.68	39.15

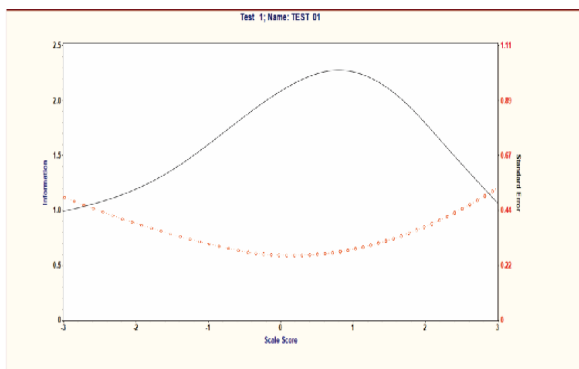
Pada tabel diatas dapat dipahami bahwa respon pada kategori skor dapat mengindikasikan tingkat kesulitan pada sebaran butir yang sesuai dengan aspek, subaspek, dan materi Biologi. Jika persentase pada kategori 1 paling besar dibandingkan dengan kategori lain dapat dikatakan bahwa pada matrik yang bersesuaian mempunyai tingkat kesulitan yang relatif tinggi, sedangkan sebaliknya jika persentase respon terbesar pada kategori 4 berarti mempunyai tingkat kesulitan yang relatif rendah. Hal tersebut dikarenakan skor 1 adalah skor terendah dalam satu butir dan skor 4 adalah skor tertinggi pada butir.

4. Reliabilitas Tes Pada Uji Coba

Berdasarkan hasil analisis dengan menggunakan program aplikasi QUEST diperoleh reliabilitas instrumen tes sebesar 0,86. Nilai tersebut dapat dikategorikan bahwa instrumen tes berpikir kritis mempunyai reliabilitas yang tinggi.

5. Kemampuan yang Sesuai Terhadap Instrumen yang Dikembangkan

Fungsi informasi berdasarkan hasil analisis pada Gambar 3 diperoleh bahwa instrumen tes berpikir kritis dapat digunakan untuk mengukur kemampuan peserta didik dengan rentang tingkat kemampuan sebesar -2,7 sampai dengan 2,9.



Gambar 3 : Fungsi Informasi dan Standar Error Pengukuran

KESIMPULAN

Berdasarkan hasil yang diperoleh dapat disimpulkan bahwa instrumen tes berpikir kritis memenuhi syarat untuk digunakan dalam mengukur kemampuan berpikir kritis peserta didik SMA Negeri Kelas X MIPA pada mata pelajaran Biologi Semester 1 dengan karakteristik Instrumen terdiri dari 2 set tes (Tes A dan Tes B) dengan

jumlah butir 25 dengan 5 butir anchor. Beberapa karakteristik yang dapat dijelaskan pada instrumen yang dikembangkan adalah :

- Instrument tes memenuhi syarat validitas isi expert judgment dan memperoleh bukti empiris kecocokan model (goodness of fit tes) pada model PCM (Partial Credit Model) berdasarkan skor politomus empat kategori.
- Instrument tes mempunyai tingkat kesulitan yang baik dengan rentang nilai -2,00 dan 2,00.
- Instrument tes mempunyai nilai reliabilitas yang tinggi yaitu sebesar 0,86 sehingga instrumen tes berpikir kritis memenuhi syarat sebagai alat ukur yang baik.
- Instrument tes dapat digunakan untuk mengukur kemampuan berpikir kritis dengan rentang kemampuan sebesar -3,7 sampai dengan 2,90.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan dalam bentuk yang pendek, ditujukan kepada sponsor riset atau pihak yang tidak bisa disebutkan dalam bagian penulis.

REFERENSI DAN SITASI

- Adams, R.J dan Khoo, S.T. (1996). *Quest : The interactive test analysis system version 2.1*. Victoria: The Australian Council for Educational Research.
- Akbar, M.N, Firman, H and Rusyati, L. 2017. *Developing Science Virtual Test to Measure Students' Critical Thinking on Living Things and Environmental Sustainability Theme*. IOP Conf. Series: Journal of Physics.
- Anisa. Aries. 2017. *Meningkatkan keterampilan berpikir kritis peserta didik melalui pembelajaran ipa berbasis potensi lokal jepara*. Yogyakarta : UNY/Jurnal Inovasi Pendidikan IPA.
- Ennis, 1986. A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44-48.
- Browne, M.W., & Cudeck, R. (1993). *Alternative ways of assessing model fit*. In K. A. Bollen & J.S Long (Eds), *Testing Structural Equatin Models*. Newbury Park, CA: Sage, 136-162 (Electronic Version).
- Hambleton, Ronald K. 1991. *Fundamentals Of Item Respon Theory*. London : New Dehli.
- Hartini dan Sukardjo. 2015. *Pengembangan Higher Order Thinking Multiple Choice Test Untuk Mengukur Keterampilan Berpikir Kritis IPA Kelas VII SMP/MTS*. Yogyakarta : UNY/Jurnal Inovasi Pendidikan IPA.
- Huber, C.H and Kuncel, N. R. 2016. *Does College Teach Critical Thinking? A Meta-Analysis*. Review of Educational Research June, Vol. 86, No. 2, pp. 431 – 468.

- Istiyono, Edy. 2013. *Pengukuran Kemampuan Berpikir Tingkat Tinggi Fisika Peserta Didik SMA di DIY*. UNY. Eprint.uny.ac.id.
- Lane, David and Oswald, Frederick L. 2016. *Do 45% of College Students Lack Critical Thinking Skills? Revisiting a Central Conclusion of Academically Adrift*. Educational Measurement :Issues and Practice Fall, Vol.35, No.3, pp.23–25.
- McPeck, J. 1981. *Critical Thinking And Education*. New York: St. Martin's.
- Mardapi, Djemari. 2016. *Pengukuran, Penialian dan Evaluasi Pendidikan*. Yogyakarta : Nuha Medika. Hlm : 12
- Moore, Betsy and Stanley, Tood. 2010. *Critical Thinking and formative assesment: increasing the rugor in your classromm*. Eye Education .
- Palm, Torulf. 2008. *Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature*. Practical Assesment, Research & Evaluation. Volume 13, Number 4.
- Peraturan Menteri Pendidikan dan Kebudayaan Tahun 2016 Nomor 23 Tentang Standar Penilaian Pendidikan
- Retnawati, Heri. 2014. *Teori Respon Butir dan Penerapannya*.Yogyakarta : Parama Publishing Hlm. 32-34.
- Stephen P. 1988. *Controlling for background belief when developing Multiple-choice CT Test*. Educational Measurement : Issues and Practice, Fall.
- Sudjana, Nana. 2013. *Penilaian Hasil Proses Belajar Mengajar*. Bandung : PT. Remaja Rosdakarya. Hlm. 3
- Travis, T. York. 2015. *Defining And Measuring Academic Succes*. The Pennsylvania State University. Practical Assessment, Research & Evaluation, Vol 20, No 5.
- Towle, A. 1989. *Modern Biology*. Austin: Holt, Rinehart and Winston. Hibbard, K.M. t.t. *Performance Assessment in The Science Classroom*. New York: McGraw-Hill Companies. Hlm 16-31
- Wilson, Suzanne M and Peterson, Penelope L. 2006. *Theories of Learning and Teaching What Do They Mean for Educators?*. National Education Association. USA : Washington, DC.